Catalogue no. 92F0138MIE — No. 001
ISSN: 1481-174X
ISBN: 978-0-662-44674-3

Research Paper

Geography Working Paper Series

How Postal Codes Map to Geographic Areas

by Kaveri Mechanda and Henry Puderer

Geography Division
3rd floor, Jean Talon Building, Ottawa, K1A 0T6

Telephone: 613-951-0437 Fax: 613-951-0569

STATISTICS CANADA
GEO•GÉO
STATISTIQUE CANADA

Statistics Canada Statistique Canada

Canada

Geography Working Paper Series

How Postal Codes Map to Geographic Areas

By

Kaveri Mechanda and Henry Puderer

92F0138MIE No. 001
ISSN: 1481-174X
ISBN: 978-0-662-44674-3

Geography Division, Statistics Canada
3rd floor, Jean Tallon Building, Ottawa, K1A 0T6

How to obtain more information :
Telephone: 613 951-0437
Fax: 613 951-0569
E-Mail inquiries: geohelp@statcan.ca

February 2007

Published by authority of the Minister responsible for Statistics Canada

© Minister of Industry, 2007

All rights reserved. The content of this electronic publication may be reproduced, in whole or in part, and by any means, without further permission from Statistics Canada, subject to the following conditions: that it be done solely for the purposes of private study, research, criticism, review or newspaper summary, and/or for non-commercial purposes; and that Statistics Canada be fully acknowledged as follows: Source (or "Adapted from", if appropriate): Statistics Canada, year of publication, name of product, catalogue number, volume and issue numbers, reference period and page(s). Otherwise, no part of this publication may be reproduced, stored in a retrieval system or transmitted in any form, by any means—electronic, mechanical or photocopy—or for any purposes without prior written permission of Licensing Services, Client Services Division, Statistics Canada, Ottawa, Ontario, Canada K1A 0T6.

Acknowledgements

The authors would like to thank Belinda Ha for her analysis and tabulations in SAS, Robert Parenteau, Sandra Porter and Joe Kresovic for their editorial help as well as Cathy Caldwell for implementing the automated geocoding.

Geography Working Paper Series

The Geography Working Paper Series is intended to stimulate discussion on a variety of topics covering conceptual, methodological or technical work to support the development and dissemination of the division's data, products and services. Readers of the series are encouraged to contact the Geography Division with comments and suggestions.

La version française de cette publication est disponible sur demande (n° 92F0138MIF au catalogue).

Table of contents

Overview **4**

Abstract **5**

Introduction **6**

1.0 Postal code typology and location **8**

1.1 The postal code and location 8
1.2 Address and delivery information from CPC files 9
1.3 The Address Lookup File 10
1.4 Files for locating postal delivery installations 17
1.5 PCType and delivery mode type from the Delivery Mode File 17
1.6 Considerations for geocoding based on the CPC data 18

2.0 Postal code users and geography **19**

2.1 Persons and the postal code reported in the census 20
2.2 Identifying the search area for geocoding 22
2.3 Geocoding postal codes directly using the census data 23
2.4 Recommendations for geocoding based on the census response 25

3.0 Geocoding rules and measures of quality **26**

3.1 Service areas and search areas 26
3.2 Geocoding postal codes by PCType 27
3.3 Measures of quality 28

Conclusion **34**

Overview

Statistics Canada creates files that provide the link between postal codes and the geographic areas by which it disseminates statistical data. By linking postal codes to the Statistics Canada geographic areas, Statistics Canada facilitates the extraction and subsequent aggregation of data for selected geographic areas from files available to users. Users can then take data from Statistics Canada for their areas and tabulate this with other data for these same areas to create a combined statistical profile for these areas.

An issue has been the methodology used by Statistics Canada to establish the linkage of postal codes to geographic areas. In order to address this issue, Statistics Canada decided to create a conceptual framework on which to base the rules for linking postal codes and Statistics Canada's geographic areas. This working paper presents the conceptual framework and intends to accomplish the following:

- study the postal code concepts and descriptors used by Canada Post Corporation in terms of their typology and occurrence
- examine how postal codes are used and construct postal code service areas based on the usage
- establish business rules governing the geocoding or linkage of postal codes to standard geographic areas and provide indicators of data quality and other metadata

The methodology described in this paper will be the basis for linking postal codes to the 2006 Census geographic areas. This paper is presented for feedback from users of Statistics Canada's postal code related products.

The link between postal codes and Statistics Canada's geographic areas is currently provided in the Postal Code Conversion File (PCCF) and the Postal Code Federal Riding File (PCFRF). The expectation is that the geocoding rules and recommendations for quality indicators can be used to create these products. These new rules have been tested on the linkages to 2001 Census *geographic areas* and have resulted in revision of over 40,000 records or about 2% of the records in the PCCF. The testing done to date indicates that about 90% of the postal codes can be geocoded with the new geocoding rules established in this working paper. More research is needed on geocoding the other 10% of the postal codes – the more recent data from the 2006 Census is expected to aid in geocoding some of these postal codes. The PCCF and PCFRF are now planned to have their postal codes geocoded to the 2006 Census geographic areas using the new geocoding rules.

Abstract

Statistics Canada creates files that provide the link between postal codes and the geographic areas by which it disseminates statistical data. By linking postal codes to the Statistics Canada geographic areas, Statistics Canada facilitates the extraction and subsequent aggregation of data for selected geographic areas from files available to users. Users can then take data from Statistics Canada for their areas and tabulate this with other data for these same areas to create a combined statistical profile for these areas.

An issue has been the methodology used by Statistics Canada to establish the linkage of postal codes to geographic areas. In order to address this issue, Statistics Canada decided to create a conceptual framework on which to base the rules for linking postal codes and Statistics Canada's geographic areas. This working paper presents the conceptual framework and the geocoding rules. The methodology described in this paper will be the basis for linking postal codes to the 2006 Census geographic areas. This paper is presented for feedback from users of Statistics Canada's postal code related products.

Introduction

This paper has two objectives: (1) to establish a conceptual framework articulating the relationship between postal codes and geographic areas; and, (2) to subsequently, establish rules for linking postal codes to *standard geographic areas.*[1]

Postal codes are managed by Canada Post Corporation (CPC) for the efficient sorting and delivery of mail. They are not created as units for the analysis or mapping of population, business or dwelling characteristics. The postal code products created by Statistics Canada provide the link between postal codes and Statistics Canada's standard geographic areas to allow for these various uses of the data. Generally, the uses of the postal code products can be characterized as follows:

1. Association of postal codes to standard geographic areas (for data retrieval, mapping, profiling and/or analysis);
2. Selection of ***statistical units*** by standard geographic areas for sampling, market research etc.

The association between the postal code and standard geographic areas is currently provided in the following products: (1) Postal Code Conversion File (PCCF) and (2) Postal Code Federal Riding File (PCFRF). The PCCF provides the linkage between the postal code and most standard geographic areas including the detailed entities: *block-face, dissemination block* and *dissemination area.* The PCCF is the primary product that meets the demands of the above-mentioned applications.

Statistical unit

A Statistical unit is the unit of observation or measurement for which data are collected or derived. The basic statistical unit for social statistics is the person, which is the unit of analysis for most social statistics programs. Other units used in social statistics include the *dwelling* and the *household.*

Over the years there have been a number of issues related to PCCF quality which have led to the questioning of the fundamental concepts underlying how the postal codes are linked to standard geographic areas. These include issues such as the appropriate use of data in linking postal codes, the precision and accuracy of the linkages and the assignment of *single link indicators.* Investigations of these PCCF quality issues have increased the already high cost, to Statistics Canada, of geocoding postal codes. These issues have highlighted a need to review and establish a conceptual framework articulating the relationship between postal codes and standard geographic areas in order to re-establish the business rules for linking postal codes to standard geographic areas and subsequent product development.

The following situations lead to difficulties in linking postal codes to geographic areas:

- There was no explicit relationship established between the postal code's municipality/community name in the CPC files and the standard geographic areas. This relationship has been on occasion subject to errors in interpretation yielding incorrect linkages.
- Since the typical geographic footprint of different types of postal codes was not well understood, the postal codes were sometimes linked to too many (or too few) geographic areas resulting in overcoverage (or undercoverage).
- Sometimes the overcoverage mentioned above resulted in a postal code being linked to too many smaller geographic areas (such as blocks) implying a level of precision that was not appropriate.

1. The terms in *italics* are defined in the glossary.

The premise of this working paper is that better use of the postal code information collected in the census as well as better use of the information in CPC files can address the issues described above as well as lead to more precise links between postal codes and standard geographic areas.

In order to improve the linkage, rules need to be developed to geocode postal codes to Statistics Canada's standard geographic areas. *Geocoding* can be defined as the process of assigning geographic identifiers (codes) to map features and data records. The resulting geocodes permit data to be linked geographically. For example, one way to geocode a *civic address* in Canada would be to start by locating the general area and then "zero-in" to the specific building with that civic address. Since the same street name or *street address* number may be used in many parts of Canada, being able to limit the searching for a street name and street address number combination to a particular place within Canada is essential. The term "*search area*" will be used in this paper to denote the geographic area within which the geocoding can be done using road names and civic addresses. The geocoding process for a civic address can then be described as follows:

Establish search area

↓

Find search area in Road Network File

↓

Find street within search area on Road Network File

↓

Find address within Address Range on street

↓

Establish Geocodes

(The geocodes used are *block-faces*, *dissemination blocks* and *dissemination areas*)

The geocoding process described above is the general model that will be used to establish business rules for geocoding postal codes. The geocoding process for postal codes ultimately creates the link between the postal codes and the geocodes of block-face, dissemination block and dissemination area. The link between postal codes and all other standard geographic areas is based on this initial geocoding.

The intents of this working paper are the following:

1. Study the postal code concepts and descriptors used by Canada Post Corporation in terms of their typology and occurrence as the first step in understanding how postal codes are made available for use by Canada Post Corporation. This is done in section 1.
2. Examine how postal codes are used and construct *postal code service areas* based on the usage patterns in order that these service areas may be used as search areas for *geocoding*. This is done in section 2.
3. Establish business rules governing the geocoding or linkage of postal codes to standard geographic areas and provide indicators of data quality and other metadata. This is done in section 3.

The expectation is that the geocoding rules and recommendations for quality indicators can be used to create the products.

1.0 Postal code typology and location

Examining the structure and typology of the postal codes is the first step in understanding how postal codes can be associated with standard geographic areas. Some of the pertinent definitions as well as the types and occurrences are examined in this section. This provides the basis for the business rules for geocoding postal codes (see section 3.0).

1.1 The postal code and location

The postal code is maintained by the CPC. It was designed to help sort mail rapidly and make more efficient the delivery of letters, parcels and other mail by CPC. A postal code may be linked to different types of points of delivery including residential mail boxes, super boxes as well as post office mail boxes. (See www.canadapost.ca).

The characters that form the postal code are generally representative of the intended service (delivery) of the postal code. Canada Post defines a postal code as follows:

> *A six-character alphanumeric combination (ANA NAN) assigned to one or more postal addresses. The postal code is an integral part of every postal address in Canada and is required for the mechanised processing of mail. Postal codes are also used to identify the various CPC processing facilities and delivery installations* (Canada Post Corporation, 2005a, p. 24).

The first three characters of the postal code (alphanumeric combination of "ANA") form the forward sortation area (FSA). This is defined as follows:

> *The forward sortation area (FSA) represents a specific area within a major geographic region or province. The forward sortation area provides the basis for the primary sorting or forward mail (Canada Post Corporation, 2006, Section B, Chapter 3, p. 9).*

The last three characters of the code are referred to as the Local Delivery Unit (LDU). These allow for the creation of individual postal codes serviced by postal installations within the FSA.

Each of the first letters of the postal code is reserved for a particular province/territory of the country (see Appendix A for more details.) Specifically, based on examining the CPC files, the first character is indicative of the province/territory of the delivery installation from which service for the postal code is provided. The second character of the postal code is indicative of the coverage of the postal code. A postal code with a "0" in the second character is classified as "rural" and all other postal codes are considered as "urban" by Canada Post.[2] However, this does not signify that the postal code applies to "rural areas" or "urban areas" based on the current census population.[3,4] We can term these as postal codes servicing *rural delivery areas* and *urban delivery areas* to avoid confusion with the Statistics Canada usage of the terms *rural areas* and *urban areas*. For example, New Brunswick does not contain any rural delivery areas although Statistics Canada does classify much of New Brunswick as rural. In the case of postal codes associated with rural delivery areas, the Canada Postal Guide states that: "*...the last three characters together with the forward sortation area identify a specific rural community*" (Canada Post Corporation, 2006, Section B, Chapter 3, p.10). The communities need to be located on a map so that they can be geocoded to standard geographic areas.

Maps are available from CPC that show the organizational area represented by an FSA. However, people living in one province may access mail using a postal code associated with another province, and similarly people living in one FSA may use the postal code of another FSA. This means that the area serviced by the postal code or the FSA may be different from the administrative boundary created by Canada Post. Further to this, the scale of these maps (1:50,000 to 1:12,000,000) makes the FSA boundaries difficult to relate to the more detailed road network in Statistics Canada's Geographic Frame. While these maps are useful in orienting the FSA within the province they neither provide the extent of the FSA with reference to a publicly available road network nor the more detailed information needed for geocoding the postal code. However, letter carrier walk (LCW) maps recently available on the internet (generally for urban delivery areas) do provide FSA boundaries in the context of road network. These can be more helpful than the FSA maps in determining the boundaries of many FSAs. However, they do not delineate the area serviced by a postal code.

The postal code by itself is not sufficient to obtain the exact location and coverage of the service area of the postal code. The FSA can be used to locate the general area serviced by most (but not all) of the associated postal codes. More information from other sources is needed for associating the postal code with specific standard geographic areas.

1.2 Address and delivery information from CPC files

CPC provides information on the postal codes in terms of the different aspects of mail delivery. This is relevant to how these postal codes map geographically and how they are accessed by the people and institutions that use the mail system. The typologies and location information described here are derived from data provided on a monthly basis by CPC to Statistics Canada in the following files:

---

2. The CPC classification is considered to be a component of mail preparation. The classification is typed based on the 2nd character of the postal code as follows:
   - an urban postal code: numerals 1 to 9 (E2J). Urban postal codes are generally serviced by letter carrier or community mailboxes; or
   - a rural postal code: numeral 0 (zero) (A0A). Rural postal codes are serviced by rural route drivers and/or postal outlets.

3. In this working paper, "census" refers to the Census of Population.
4. Statistics Canada defines urban and rural areas, based on the Census of Population reported for dissemination blocks, as follows:

   An urban area has a minimum population concentration of 1,000 persons and a population density of at least 400 persons per square kilometre, based on the current census population count. All territory outside urban areas is classified as rural. Taken together, urban and rural areas cover all of Canada.

1. Address Lookup File
   This file contains all the active postal codes and all the valid mailing addresses in the country (Canada Post Corporation, 2005a). An updated version of this file is available on a monthly basis from CPC and it contains all the postal codes that are considered "active" for mailing at that time. About 97% of the postal codes in the file are associated with delivery to a civic address.[5] Such postal codes can be associated with more than one address range record in the file. These records include a community name maintained by CPC, a street name and a civic address range. (The various types of records included in the Address Lookup File and the *CPC community* name are described in section 1.3.)
2. Householder File
   This file is designed to provide information on the number of dwellings and businesses served by delivery installations. It contains all rural and urban delivery installations in Canada (Canada Post Corporation, 2005b). It also contains the addresses of some of these installations. These installation addresses can be useful for geocoding postal codes to their installations. (The use of the Householder File is discussed in section 1.4.)
3. Postal Code Delivery Mode File
   This file contains postal codes and their related delivery modes. (Canada Post Corporation, 2005c). It indicates whether the delivery is made to a civic address range, to a business building or to a *post office box*. (The delivery mode types are discussed in more detail in section 1.5.)

These are the three most relevant files associated with the geocoding of postal codes. Of these the Address Lookup file, containing the address information associated with the vast majority of the postal codes, is the most useful for geocoding.

1.3 The Address Lookup File

The Address Lookup File shows the complete picture in terms of active postal codes (including updates and about 500 to 2500 new postal codes introduced or "birthed" each month). Of the 1,144,108 records in the June 2006 Address Lookup file, 1,133,814 were "active" or valid for mailing. The other postal codes in the file are those that are "retired" that month. Statistics Canada uses the term "retired postal code" to describe all postal codes that were active at some time but, are no longer active. Although Canada Post only considers the active postal codes as valid for mailing, retired postal codes may be used for some time after they are retired.

The CPC community name in the Address Lookup File

Identification of the municipality serviced by a postal code is an important initial step in "zeroing in" on where the postal code should be geocoded. The 1st character of the postal code (indicating the province or territory) in combination with the CPC community name does in some (but not all) cases point to the current municipality or *census subdivision* (CSD) to which the postal code can be geocoded. When the CSD can be located, civic address information could be used to geocode to a street address range.

The term "municipality" as in the Address Lookup File does not necessarily constitute the legally defined municipality at the time when Canada Post publishes its data. These names serve the purpose of preserving locally used names to prevent confusion in terms of the area serviced. The area serviced does not necessarily coincide with legally constituted municipalities of the same or similar name. For example, when municipalities are amalgamated, Canada Post preserves the older (non-amalgamated) names until it can establish that different streets are uniquely named in the newly amalgamated municipality. The FSA is also defined in relation to a municipality as follows:

5. The other 3% of the postal codes are not associated with civic addresses but, with post office boxes and postal service routes.

The FSA comprises the first three characters of a postal code. The FSA qualifies the area within a municipality that is permitted to use the valid alternate municipality name (Canada Post Corporation, 2005a, p 23).

These municipality names are often names of places no longer in existence or commonly used names associated with localities. To avoid confusion this name will be referred to hereafter as the "CPC community".

Comparison of municipal names in use in January 2001 illustrates this point. Statistics Canada's CSDs contain all the legally defined municipalities existing on January 1st of the census year (2001). A comparison of the CPC community names in existence shows that less than 40 % matched the CSD. The following table illustrates the matches of CPC community name with the 2001 CSD names[6].

Table 1 CPC community names and CSD names

Address Lookup File		Match to 2001 Census CSDs	
Vintage*	Number of *CPC* Communities		
		number	%
November 2000	6,983	2,689	38.5
June 2001	7,012	2,697	38.5
December 2003	6,987	2,708	38.8
December 2004	6,970	2,703	38.8
June 2005	6,961	2,701	38.8
October 2005	6,958	2,707	38.9

*The vintage dates reflect postal codes made/kept active for that month and year.
Data sources: Canada Post Corporation, November 2000 - October 2005, Address Lookup File, and Statistics Canada, 2001 Census of Population.
Table source: Statistics Canada, 2007, How Postal Codes Map to Geographic Areas, Catalogue number 92F0138MWE2007001.

The match rate in the above table is about 40% despite some processing (standardization) to increase the match rate. Examination of the June 2001 Address Lookup File shows changes occurring as a result of amalgamation. For example, both Nepean and Ottawa are maintained as CPC community names although Nepean no longer exists as a legal municipality as a result of amalgamation. The street names and the CPC community names appear to be updated in the Address Lookup File to reflect the changes at the municipal level once CPC resolves the uniqueness of the address. This may explain the slight increase in match rates over four years. However, for geocoding purposes, a correspondence file between CSDs and the CPC community names is needed so that the geocoding can be narrowed down to the CSD first and then to the street address within it.[7]

Unfortunately even where the CSD names match, more information is needed to define the area within which the street address information can be uniquely matched. More than one street can

6. This match rate is based on matching unique CSD names within a province/territory with the *CPC community* names of postal codes of a province/territory (based on the first letter of the postal code). The matching was done after standardising the names to all capitals with no accents or special characters. The figures in the table are in fact a slight overestimate of the match rate since the CSD type could not be factored in the matching and is not in the denominator of the percentage calculation. (CSDs with the same name and a different type (town, reserve, municipal district etc.) such as Yarmouth Town and Yarmouth Municipal District both in Nova Scotia constitute entirely different CSDs. Please see the glossary on p. 40 for a definition of CSD Type.)
7. Finding the particular *service area* of a postal code is essential to establishing an automated geocoding process. However, in a manual geocoding process, the options exist to look-up various maps and even consult municipalities.

occasionally have the same name within a CSD. The same postal code may sometimes be assigned to multiple CSDs. The possibility of geocoding within multiple *CPC communities* should be considered when geocoding.

The CPC community name can be a useful tool in "zeroing in" to the municipalities within which the street address may be found. Sometimes a correspondence can be established between the CSD and the CPC community name based on province and the similarity of the name. But, because of the differences in vintage and purpose of the CSD name and the CPC community name, this correspondence is not always possible, and even when established can not be done with absolute confidence. Supplementary information is needed in establishing the area within which the street address may be found.

Postal code types in the Address Lookup File

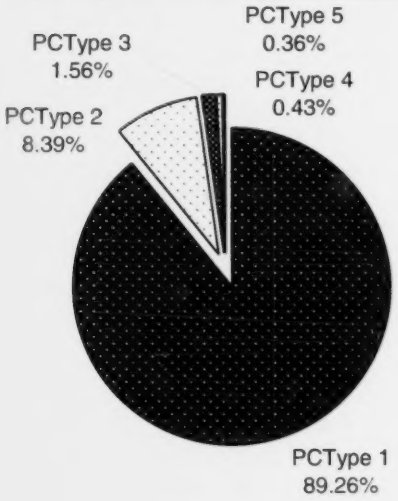
Once the municipality or general area serviced by the postal code is established, the specific service area of the postal code needs to be understood before geocoding. The record types in the Address Lookup File are essential in establishing the type of service area for the postal code. Canada Post Corporation designates five types of records in the Address Lookup File.[8] These are associated with five different types of delivery. These records are typed and named as "PCType" in the PCCF file. The typology is as follows:[9]

- PCType 1, Street address
  This record contains postal codes for civic addresses.
- PCType 2, Street address served by route or *general delivery*
  This record contains postal codes for street address served by route service or general delivery from a postal delivery installation – often occurring in the fringes of urban areas.
- PCType 3, Lock box address
  This record contains postal codes for a *lock box* or a range of lock box numbers within a postal delivery installation.
- PCType 4, Route service address
  This record contains the postal code for a *rural route*, *mobile service* or suburban service route served by a postal delivery installation.
- PCType 5, General delivery address
  This record contains postal codes for general delivery service at a postal delivery installation.

A postal code can be associated with more than one PCType. Each record in the June 2006 Address Lookup File has a postal code and a specific PCType. Figure 1 shows that the vast majority of the records are of PCType 1 or 2:

8. This information is based on an analysis of Postal Code data and information gleaned from the Canada Post Corporation documents on the CPC website www.canadapost.ca. as well as a telephone conversation with a Canada Post representative on November 17th, 2004.
9. This typology list is an amalgamation of information provided for records described in "Postal Code Address Data Technical Specifications" (Canada Post Corporation, 2005a, p.5-10).

Figure 1 June 2006 Address Lookup File records by PCType



Data sources: Canada Post Corporation, June 2006, Address Lookup File.

The service areas of records with PCType of 1 and 2 are defined in terms of civic addresses. Basically civic (street) addresses are those that are found on dwellings or buildings. As can be seen above, the PCTypes with civic addresses constitute about 98 % of the records in the Address Lookup File in June 2006. Most urban block-faces and many residential streets in urban areas typically have only one civic addressed postal code associated with them. While there may be difficulties in finding the CPC community and locating the standard geographic area*(s)* for a postal code, the service areas of these postal codes can be considered as completely defined by the address information provided by Canada Post. Here, geocoding is a simple matter *if* the road name and address information can be matched to the same vintage data in Statistics Canada's Geographic Frame.

No civic address information for the service area is provided for PCTypes 3, 4 and 5. The service areas of these postal codes are not directly defined in the Address Lookup File. For PCType 3, the Address Lookup File provides lock box or post office box address ranges within a postal delivery installation. All postal codes can be linked back to the installation from which they are serviced.[10] At a minimum, the postal codes of PCTypes 3, 4 and 5 can be coded to the CSD where the postal delivery installation is located (if the postal delivery installation address can be geocoded). For PCType 4, it is possible to sometimes obtain the names of the routes serviced. These include delivery along well-defined roads in settled rural areas, deliveries to industrial parks and to group mail boxes. For PCType 5 the only information is the CPC community name of the area serviced. In all cases, the service area is not necessarily where the postal delivery installation is located and the service area is not defined in terms of civic address ranges. In short, the postal codes with PCTypes 3, 4 or 5 (unlike PCTypes of 1 and 2) do not have a precise geographic footprint that permits direct and complete association with Statistics Canada's geographic files. Alternate data sources depicting usage patterns need to be examined to decide on how to allocate service areas for these postal codes and how to geocode these postal codes based on that information.

10. Information from the Address Lookup File can be used to find the delivery installation in the Householder File. The civic address of the delivery installation is often but not always available in the Householder File. Point locations of the delivery installations are also available from Canada Post Corporation.

With increased urbanisation and civic addressing, the proportion of PCType 2 postal codes has increased slightly. Table 2 illustrates the changes over time:

Table 2 Changes in postal code PCTypes combinations in the Address Lookup File

PCType	June 2001				June 2006			
	Urban delivery area		Rural delivery area		Urban delivery area		Rural delivery area	
	number	percent	number	percent	number	percent	number	percent
1	739,624	98.2	0	0.0	786,435	98.2	0	0
2	0	0.0	0	0.0	0	0	0	0
3	12,039	1.6	817	15.7	12,865	1.6	841	16.1
4	250	0.0	156	3.0	249	0	136	2.6
5	626	0.1	615	11.8	647	0.1	457	8.8
1, 3	46	0.0	0	0.0	46	0	0	0
2, 4	864	0.1	290	5.6	760	0.1	443	8.5
2, 4, 5	0	0.0	106	2.0	0	0	161	3.1
2, 3, 5	0	0.0	11	0.2	0	0	39	0.8
2, 3, 4	0	0.0	88	1.7	0	0	156	3
2, 3, 4, 5	0	0.0	482	9.3	0	0	677	13
2, 5	0	0.0	6	0.1	0	0	17	0.3
3, 5	0	0.0	1,203	23.1	0	0	1,259	24.2
3, 4	0	0.0	288	5.5	0	0	212	4.1
3, 4, 5	0	0.0	864	16.6	0	0	631	12.1
4, 5	0	0.0	283	5.4	0	0	185	3.6
Total	**753,449**	**100**	**5,209**	**100.0**	**801,002**	**100**	**5,214**	**100**

0 true zero or a value rounded to zero

Note: The data units in this table are records grouped by unique postal code PCType combinations.

Data sources: Canada Post Corporation, June 2001 and June 2006, Address Lookup File.

Table source: Statistics Canada, 2007, How Postal Codes Map to Geographic Areas, Catalogue number 92F0138MWE2007001.

While the percentage of each type of urban postal codes has remained consistent, the percentage of postal codes that include PCType 2 has had a small increase in rural delivery areas. This slight gain in the proportion of postal codes with civic address range records could lead to a slight increase in the proportion of postal codes that can be geocoded to a civic address range.

As can be seen, a postal code could have records associated with more than one PCType. This generally occurs in the areas that Canada Post considers to be "rural". PCType 2 does not occur in isolation – it occurs with PCTypes 3, 4, and 5. The proportion of postal codes that have PCType 2 records continues to increase. This may be an indication of an increase in civic addressing, particularly where there has been route service (PCType 4). For example, postal code K7V3Z8 is associated with a postal code of route service (PCType 4). However, it also has records of PCType 2 with civic addresses associated with it. Very rural areas tend to have only PCType 5 and/or 3 postal codes.

PCType 1 occurs in urban areas and tends to occur in isolation except in a few cases with PCType 3. An example of this type of situation is the postal code K1A 9Z6.[11] This is associated with a post office box and a civic address. The civic address record (PCType 1) is used to supplement the information in the record with the delivery installation address (PCType 3) for the

11. This postal code with PCType=1 and PCType=3 was found on the December 2004 version of the Address Lookup File from Canada Post Corporation. This was also confirmed on the Internet lookup tool at www.canadapost.ca (accessed May 30, 2005).

same postal code. [12] Effectively, these postal codes have two PCTypes but refer to one physical building (or business) that receives mail.

The service area of a postal code can be considered to be the area where persons live and have access to mail for that postal code. For postal codes of PCType 1 and 2 that would include the dwellings and buildings with mail delivery based on their civic addresses. For postal codes with PCType 4, all of the areas that receive mail by route service would be included. For postal codes with PCType 3 and 5 that would encompass the locations of dwellings and buildings where persons typically access mail delivery through the associated delivery installation. The service area of an individual postal code would include that of each of its PCTypes. Examination of the PCType combinations associated with each postal code can help in understanding if and how the coverage, of these service areas, overlaps. This in turn can also help in deciding whether certain records are more representative of the geographic footprint of the postal code.

Some PCTypes may be considered as more representative of the definitive service area for that postal code. For example, PCType 1 is more representative of a specific building/dwelling serviced than PCType 5. PCType 4 is representative of specific routes serviced whereas PCType 5 is not (based on the address information available) associated with one particular location more than any other within the CPC community. As discussed earlier in this section, where PCType 1 and 3 occur for the same postal code, the PCType 1 record shows the address of the users of the mail service. All of this suggests that the most representative records of the postal code's coverage would be as follows:

- PCType 1 rather than PCType 3 when both are present for the postal code;
- PCType 2 or 4 rather than PCType 3 or 5 when both are present for the postal code.

Both PCTypes 2 and 4 may cover the same area. The area serviced by the route service (PCType 4) may be quite specific even if it can not be located based on the address information from the Address Lookup File. This information is not sufficient to decide which of these records (PCType 2 or 4) is most representative. However, PCType 2 may be given precedence over PCType 4 in the geocoding process as being more representative since it is likely to be better geocoded (given specific address information) and also likely represents a part of the coverage of PCType 4 as a result of the progress in civic addressing. Essentially, when more than one PCType is associated with a postal code, the PCType may be used as one of the criteria to choose the most representative indicator for the spatial footprint of that postal code.

Once the CPC community has been matched to a CSD, PCTypes 1 and 2 could be geocoded based on street address information, but, the lack of civic addressing information with PCTypes 3, 4 and 5 makes them almost impossible to geocode below the CSD level in the same way. Given that the target is to map the entire footprint of the postal code, more information is needed where postal codes have PCTypes 3, 4 and 5 records. The PCTypes can also be used as one factor in determining the most representative linkage of postal code to standard geographic area. To conclude, PCType is critical to determining the nature of the service area footprint of the postal code.

12. This is based on a telephone conversation with a Canada Post Corporation representative on November 17, 2004.

Postal code records and geographic area coverage

Even a postal code that maps specifically to civic address ranges may cover more than one block-face or block. Therefore when a postal code on the Address Lookup File has more than one record associated with it, each with its own address range, this can legitimately lead to the coverage of more than one block-face (or dissemination block). (This can also happen if a single address range from the Address Lookup File spans more than one block-face or dissemination block).The tables below shows the number of records in the Address Lookup File associated with each PCType and postal code combination:

Table 3 Number of records per postal code PCType on the Address Lookup File

PCType	Records					Total records of a particular postal code			
	Total	Minimum	Maximum	Mean	Standard deviation	only 1	2 to 5	6 to 50	50 and over
	number					percentage			
June 2001									
1	912,835	1	78	1.23	0.94	72.56	21.20	6.20	0.04
2	58,414	1	512	31.63	41.11	0.12	1.18	47.02	51.68
3	16,718	1	34	1.06	0.49	91.59	7.36	1.05	0.00
4	4,996	1	10	1.36	0.84	57.77	40.23	2.00	0.00
5	4,196	1	1	1.00	0.00	100.00	0.00	0.00	0.00
Total	**997,159**	...	...	...	...	**68.68**	**19.80**	**8.45**	**3.07**
June 2006									
1	1,012,077	1	86	1.29	1.05	67.89	24.80	7.25	0.07
2	94,979	1	507	42.16	50.03	0.10	0.57	34.62	64.71
3	17,763	1	34	1.06	0.55	90.85	7.53	1.62	0.00
4	4,922	1	8	1.36	0.84	57.72	40.41	1.87	0.00
5	4,073	1	1	1.00	0.00	100.00	0.00	0.00	0.00
Total	**1,133,814**	...	...	...	...	**62.65**	**22.47**	**9.40**	**5.48**

... not applicable

Data sources: Canada Post Corporation, June 2001 and June 2006. Address Lookup File.

Table source: Statistics Canada, 2007, How Postal Codes Map to Geographic Areas, Catalogue number 92F0138MWE2007001.

Examination of the table above shows that while records of PCType 1 are related to one address range about 70% of the time, even these may sometimes cover more than 50 address-ranges (potentially 50 block-faces or dissemination blocks). In our experience in geocoding of PCType 1 and 2, the coverage is of consecutive or adjacent block-faces when a postal code has more than one civic address range record. PCType 2 appears to have the most number of defined address ranges associated with it. PCType 4 has routes associated with it and these routes may cover large areas even if there are not that many records in the Address Lookup File. PCType 3 is generally associated with one or more post offices boxes in a postal delivery installation and has only one record associated with the installation. Similarly PCType 5 refers to delivery that is available from an installation and may have only one record. Of significance here is that for PCType 1 and especially PCType 2 there may not be one single geocoded record (block-face or block) that is more representative of the postal code in terms of service provided than any other.

1.4 Files for locating postal delivery installations

PCType 3 and PCType 5 postal codes are associated with the pick-up of mail from a postal delivery installation (rather than delivery to a civic address). The civic addresses associated with the postal delivery installation are available on the Householder File. The point locations for these postal delivery installations were also recently acquired from CPC by Statistics Canada. The point locations are not always consistent with the coordinates of Statistics Canada's Road Network File. This means that where address information is available and can be matched to that in the Road Network File, the geocoding is best done with it so that the position of the installation relative to the road network information is shown as accurately as possible. When address information is not available, the post office location address can be confirmed with Canada Post. Statistics Canada could maintain a list of postal delivery installation addresses and associated dissemination blocks to permit the geocoding of postal codes. Section 2.0 examines the question of whether some postal codes should be associated with the postal delivery installation only.

1.5 PCType and delivery mode type from the Delivery Mode File

The delivery mode type (DMT) provides detail on the mail delivery service for urban delivery areas.[13] It is used to calculate the rates for mailing services (Canada Post Corporation, 2005c, p.2). The CPC description and typology is as follows:

A = Delivery to block face address
B = Delivery to an apartment building
E = Delivery to a business building
G = Delivery to a large volume receiver
H = Delivery via a rural route
J = General delivery
K = Delivery to post office box (not community mail box)
M = Delivery to a large volume receiver (post office box)
T = Delivery via a suburban service
X = Delivery via a mobile service
Z = Postal code is retired (no further delivery to this code)

The delivery mode type can be useful in illustrating whether the geocoding applies to specific building and thus may potentially be "sub" block-face. As shown in table 4, in a few cases postal code records can be associated with a specific building:

13. The DMT should not be confused with the PCType in the PCCF. The DMT has more to do with **how** the mail is delivered. The PCType of the postal code takes precedence in terms of **where** the mail is delivered. A typical case would be a large volume receiver who has a postal code with PCTypes of 3 and 1. A piece of mail with address formatted as PCType 1 (street address) would still be sent to the post office box of the postal code as long as the postal code was correct and machine readable. (Otherwise, the machine will attempt to decipher the postal code based on the address information.) This system prevents large amounts of mail being delivered by a letter carrier. The DMT is used by large volume mailers to obtain discounts based on pre-sorting mail. (This is based on a conversation with a representative of CPC, November 17, 2004.)

Table 4 PCType 1 Records from the June 2006 Address Lookup File with DMT from June 2006 Postal Code Delivery Mode file

PCTYPE	DMT	Records	Records
		number	%
1	A	972,547	85.78
1	B	20,760	1.83
1	E	10,097	0.89
1	G	8,673	0.76

Data sources: Canada Post Corporation, June 2006, Address Lookup File and Canada Post Corporation, June 2006, Postal Code Delivery Mode File.
Table source: Statistics Canada, 2007, How Postal Codes Map to Geographic Areas, Catalogue number 92F0138MWE2007001.

As can be seen, the most frequent combinations of DMT and PCType are associated with well defined address ranges (DMT = "A"). The rest of the PCType 1 are associated with apartment buildings (DMT = "B") and different types of businesses (DMT = "E" and G"). Most of the geocoding of PCType 1 can be done simply based on the address information. However, it is especially important that buildings with large populations are geocoded properly. For example, a building may represent a health facility that defines an entire block, dissemination area or census subdivision *(CSD)*. In the rare case where an entire CSD is represented by the facility, it's population could represent all of the CSD's population. For PCType 1 postal codes, the DMT helps to distinguish the type of building (dwelling, apartment, large volume receiver, etc.) that is represented by the postal code. The DMT does not add more information than the PCType for locating postal codes based on address information, although it does about the nature of the building(s) receiving the mail.

1.6 Considerations for Geocoding based on the CPC data

These considerations for geocoding follow from the examination of the CPC data as elaborated in this section. These considerations, in conjunction with the findings in the next section on how postal codes are used and reported, will be used to develop the rules for geocoding.

The considerations for developing the geocoding rules are:

- The CPC community name and the first letter of the postal code is sometimes but, not always a way to locate the CSD for geocoding. Although in theory more than 97% of the records in the Address Lookup File contain street address information, less than 40% of the unique CPC community names can be matched to the CSD structure. This makes the case for supplementary information to both confirm and complete the geocoding.
- For PCTypes 1 and 2 the CPC community name in combination with the street address information can uniquely identify the civic addresses (representing dwelling and buildings) serviced by the postal code ***if*** the area that is represented by the CPC community name can be found.
- For PCTypes 3, 4 and 5, the service area is not defined by civic addresses. Supplementary information on where these postal codes are used is needed to geocode these.
- The PCType combinations for a postal code can be used to help determine the most representative or indicator records for that postal code. However, there may be more than one address range record associated with a PCType for a postal code and these postal codes may quite legitimately be associated with more than one block-face or block.

In conclusion, supplementary information on the areas where postal code service is received is needed to both augment and validate the partial geocoding that can be done based on the CPC files alone.

2.0 Postal code users and geography

After examining the postal code data in the previous section, the key challenge remains: how to develop a search area where the postal code can be geocoded. The CPC community names are place names that do not usually match names or boundaries of the CSDs maintained in Statistics Canada's Geographic Frame. In the previous section, the discussion on the use of the CPC files for geocoding showed that supplementary information was needed for geocoding. The substantive un-answered questions were:

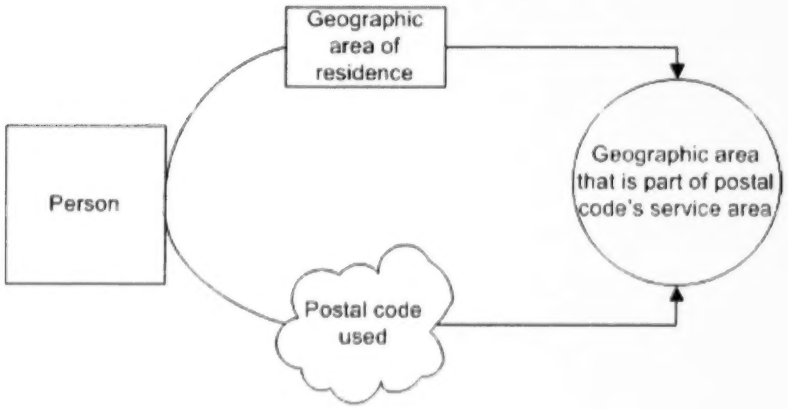
1. Since less than 40% of the unique CPC community names match those of the CSDs, how can we geocode addresses containing the remaining 60% of the CPC community names?
2. How can the service area be determined for PCTypes of 3, 4 and 5 for which we do not have civic addresses?
3. Is there supplementary information that can validate address information based geocoding?

The supplementary information being considered here is information based on the reported use of the postal code. The area where a postal code is used is referred to as the service area of the postal code. The service areas, based on the postal code reported in the Census of Population, provide the supplementary information to both enable and confirm the geocoding. This is in addition to trying to find the CSD based on the CPC community name. Improvements to this process are discussed in Appendix C. This section focuses on how persons or households use postal codes and how this provides supplementary and confirmatory information to that of the address information from CPC.

The users of the postal code are persons and businesses. But, the relationship between these statistical units and the postal codes is indirect. As discussed in the previous section, postal codes are not assigned to persons or businesses but, to buildings and dwellings.[14] The exception to this are the less than 2% of the postal code address records, with DMT of "G" or "M", which are associated large volume receivers. A data source is also needed for finding the service area in this manner. In the case of persons the relationships needed from the data source are as follows:

14. The vast majority of records in the Address Lookup File are of PCTypes 1 and 2, and these refer to delivery to a specific dwelling or building. Other PCTypes refer to services available from the delivery installation.

Figure 2 Postal code used is related to geographic area of residence



Essentially the person's relationship with the postal code and the standard geographic area can be used to make the connection between the latter two.

Even if such a data source is available for the postal code, the question remains as to which standard geographic area is related to that postal code. For example, there is a very slight risk that the standard geographic area assigned in the data source not being that of the person's home but that associated with the standard geographic area of employment. These risks need to be balanced in the context of how persons use postal codes and how businesses use postal codes.

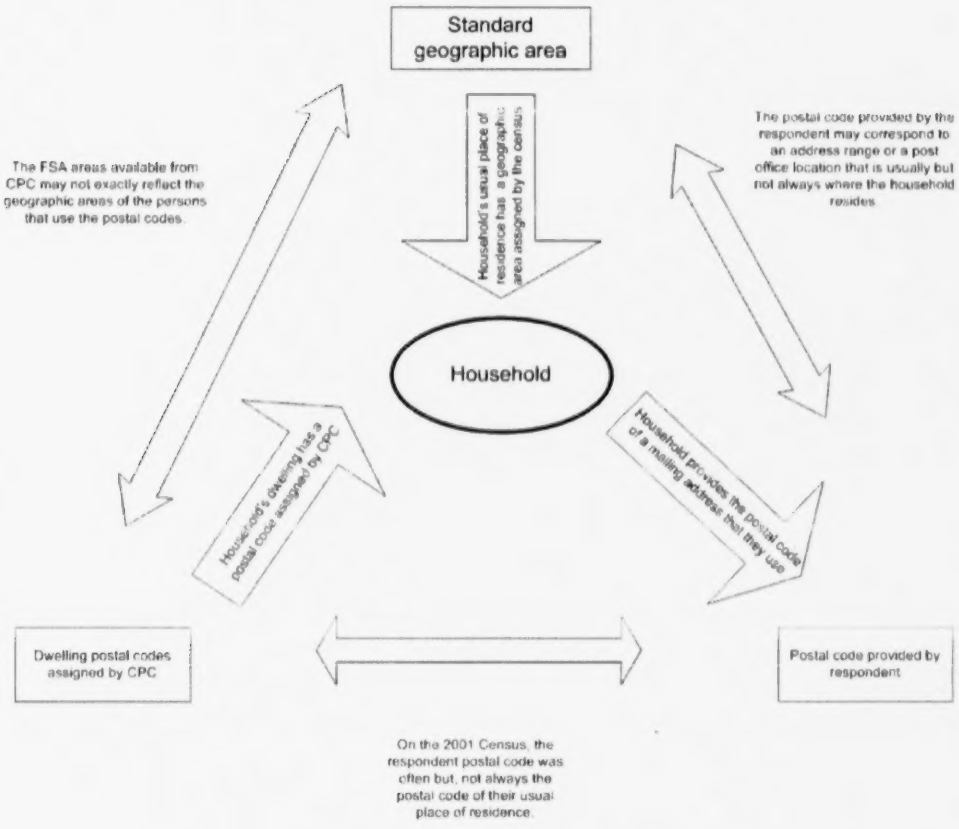
2.1 Persons and the postal code reported in the census

The one source containing a reported postal code and the standard geographic area for persons in the entirety of Canada is the Census of Population. Households in Canada are required to report (or confirm) their postal code on the census form. This respondent postal code is typically the postal code of the residence of the household. This can be used to estimate the service area of a postal code.

A household is generally defined as being composed of a person or group of persons who co-reside in, or occupy, a dwelling.

The census is the most complete survey of household postal code response data for Canada. The postal code provided by the respondent in the census and the postal code assigned to the person's dwelling by CPC is usually the same. This information can be used to find the standard geographic area related to the postal code reported in the census. The diagram below shows how these are related:

Figure 3 Household uses of postal codes


Standard geographic area
Household's usual place of residence has a geographic area assigned by the census
The postal code provided by the respondent may correspond to an address range or a post office location that is usually but not always where the household resides
The FSA areas available from CPC may not exactly reflect the geographic areas of the persons that use the postal codes.
Household
Household's dwelling has a postal code assigned by CPC
Household provides the postal code of a mailing address that they use
Dwelling postal codes assigned by CPC
Postal code provided by respondent
On the 2001 Census, the respondent postal code was often but, not always the postal code of their usual place of residence.

The census asks respondents to provide the postal code of their residence. Given that this information exists at the household and dissemination block levels, why not simply assign postal codes to dissemination blocks (and other standard geographic areas) based on the census alone? The standard geographic areas are typically containers by which analysis is done and any extraction of data using the postal code as a key can be done with these standard geographic areas. However, if standard geographic areas are to be used as the means to reliably link postal codes to other data, then the following are the reasons why the census information needs to be supplemented with geocoding:

1. Postal codes change every month, with hundreds of codes being created, changed and deleted. The census data is only available every five years and many months after Census Day. The geocoding operation allows for the timely geocoding of the most current postal codes.
2. Occasionally, more than one postal code may be used by any one person or business. A person may receive mail at their usual place of residence, an employed person who travels may use a work address and have the mail forwarded to her, a student may receive mail at his parent's address, and some persons may prefer to receive mail at a post office box. These are all examples of different addresses and different postal codes that may be used by the same person. Similarly, a business may use a postal code associated with the building or block-face where it is located, a post office box or the postal code of the mailing address of its headquarters.

3. The postal code response variable needs to be edited because there can be variations in the postal code response (as discussed earlier). Coding postal codes to a street network file on Statistics Canada's Geographic Frame allows for the confirmation that the standard geographic area is that of the postal code footprint.

The census reported postal code can be used to map service areas for both the postal code and the FSA. Thus, the best quality data can be obtained by combining the information from the census with that available from CPC.

2.2 Identifying the search area for Geocoding

Since less than 40% of the unique CPC community names match to those of the CSD, how can we geocode addresses containing the remaining 60% of the CPC community names? Locating the street names and addresses for geocoding can be done using the census data.

A test was conducted in autumn 2005 to measure the rate of geocoding based on *FSA service areas*. Since FSAs are more stable over time than postal codes, search areas were created using the FSA of the valid reported postal codes from the 2001 Census. For this test of geocoding of the June 2005 postal codes, *FSA* service areas were created using the postal codes that were valid in May 2001. Dissemination blocks were assigned to an FSA only if a valid postal code with that FSA was reported at least four times in that block.[15] The geocoding of postal codes was done within the FSA service area for that postal code's FSA. Essentially the FSA service areas were used as the search areas for the geocoding. Despite these conservative criteria, over 88% of the civic address records (PCType of 1 and 2) could be geocoded – most of them at the most detailed block-face level.

Using these search areas appears to be a productive way to do the geocoding where the *postal code* service area is designated in terms of civic addresses (PCTypes 1 and 2) and the related civic addresses and street network information is available in the Road Network File.

15. Where postal codes for another province were reported within the block, this was not considered in the geocoding. The geocoding method is explained more fully in the next section.

2.3 Geocoding postal codes directly using the census data

A method is needed for geocoding postal codes that either do not link to civic addresses (PCType 3, 4, 5) or have civic addresses that can not be found on the Road Network File. Would the census provide sufficient information to geocode these?

The 2001 Census reported postal code data can be examined to see the PCTypes of postal codes that were in use. These responses should be evaluated with some caution because some postal codes have multiple PCTypes associated with them and respondents do not state the PCType of their postal code. The postal codes are therefore examined below in terms of the combinations of PCType. The June 2001 Address Lookup File had 758,658 postal codes which were not retired. Of the 673,242 valid postal codes reported in the census, 673,179 postal codes were found to be on the June 2001 Address Lookup File.

A detailed examination of the postal codes that were ***not*** reported revealed the following:

Table 5 Postal codes that were *not* reported in the 2001 Census

	Urban delivery			Rural delivery		
	June 2001	2001 Census		June 2001	2001 Census	
PCType	Address Lookup File	Not reported	Percent not reported	Address Lookup File	Not reported	Percent not reported
	number	number	%	number	number	%
1	739,624	76,581	10.35	0	0	0
2	0	0	0	0	0	0
3	12,039	6,572	54.59	817	6	0.73
4	250	35	14.00	156	2	1.28
5	626	331	52.88	615	8	1.30
1, 3	46	3	6.52	0	0	0
2, 4	864	12	1.39	290	0	0.00
2, 4, 5	0	0	0	106	0	0.00
2, 3, 5	0	0	0	11	0	0.00
2, 3, 4	0	0	0	88	0	0.00
2, 3, 4, 5	0	0	0	482	0	0.00
2, 5	0	0	0	6	0	0.00
3, 5	0	0	0	1,203	2	0.17
3, 4	0	0	0	288	0	0.00
3, 4, 5	0	0	0	864	1	0.12
4, 5	0	0	0	283	0	0.00
Total by PC	**753,449**	**83,534**	**11.09**	**5,209**	**19**	**0.36**

0 true zero or a value rounded to zero

Data sources: Canada Post Corporation, June 2001, Address Lookup File and Statistics Canada, 2001 Census of Population.

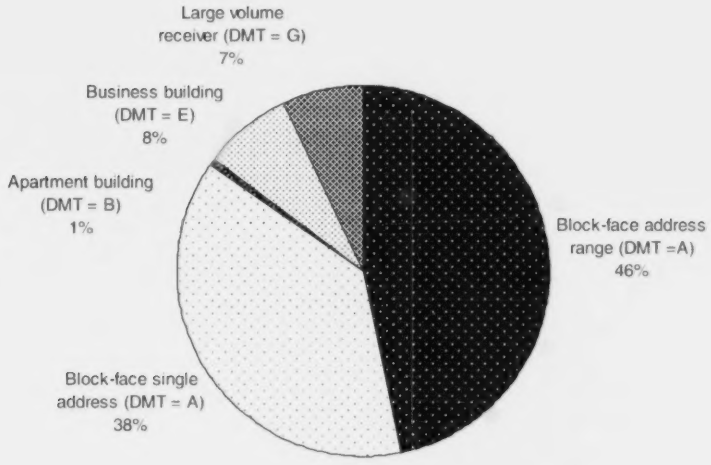
Table source: Statistics Canada, 2007, How Postal Codes Map to Geographic Areas, Catalogue number 92F0138MWE2007001.

It appears that almost every rural delivery postal code is reported. Only 19 rural delivery postal codes were not reported at all. This is less than 1% of all the rural postal codes. Since coding PCTypes 4 and 5 are particularly difficult to geocode, the use of the census responses appears to be an effective approach to geocode these in the rural areas. Comparison of the 2001 Census reported results with the 2006 Census reported results should allow us to evaluate whether these postal codes are stable over time and therefore whether the use of the census data may be sufficient for geocoding the postal codes active during the census.

Fewer urban delivery (in comparison to rural delivery) postal codes are reported. Business postal codes are understandably not generally reported in the census. This appears to be the case with about 50% of the PCType 3 postal codes. The PCType 5 postal codes may not be used as much in urban delivery areas, particularly since most people likely have delivery directly to their place of residence. Although only 10% of the PCType 1 postal codes were not reported, these still constitute a large number (76,581). These postal codes are not designated as business postal codes by PCType. However, upon closer examination, these do appear to be postal codes associated with non-residential buildings. For example K2E7L6, K2E7J5 and K2E1B6 are associated with civic addresses along Colonnade Drive in Ottawa. This area is zoned as a "general industrial zone".[16] The subset of PCType 1 postal codes that were not reported were classified further in terms of their DMT and then checked to see if they constituted delivery to a single address.

The breakdown of the PCType 1 postal codes that were not reported is shown below:

Figure 4 PCType 1 postal codes *not* reported


Large volume receiver (DMT = G) 7%
Business building (DMT = E) 8%
Apartment building (DMT = B) 1%
Block-face address range (DMT =A) 46%
Block-face single address (DMT = A) 38%

Data sources: Canada Post Corporation, June 2001, Postal Code Delivery Mode File and Statistics Canada, 2001 Census of Population.

Altogether, slightly more than half of these postal codes are associated with delivery to single address.[17] Given that more than half of these postal codes are associated with one single

16. Based on maps and information on the City of Ottawa website accessed on *Ottawa.ca*, July 5, 2006
17. This includes all of the postal codes in the chart above with DMTs of E, B and G as well as the postal codes with DMT of A that have delivery to a single address only.

address, some of them clearly designated as business address by CPC, suggests that the urban under reported PCType 1 postal codes could indeed be mostly business or institutional postal codes.

Use of the census reported postal code is likely an appropriate tool to find the service areas of the FSAs as well as that of postal codes associated with households (rather than businesses or institutions). Generally, household postal codes are reported in the census and service areas created from the reported FSA should be sufficient for geocoding the majority of postal codes.[18]

2.4 Recommendations for geocoding based on the census response

Census data is a good source for geocoding where postal codes are in use at census time. It can also be used to create search areas for geocoding postal codes with civic addresses. This baseline information can help in assigning postal codes in terms of usage to dissemination blocks, dissemination areas and CSDs.

Postal code usage can change over time both because Canada Post changes the postal codes and because users may change their reported postal codes. Further to this, it is not possible to obtain a postal code for every household in the census because the response may be illegible, invalid or blank. However, the response postal code, especially if it is frequently reported within a dissemination block can be effectively used to construct the FSA of postal code usage by block. While postal codes tend to change a little every month, the FSA service area is likely to be more stable over the five-year period between censuses.[19] The ability to geocode 2005 Canada Post data based on service areas created from the 2001 census data suggests that FSAs are indeed usually stable. The stability of the FSA service areas and as well as that of the relationship of the reported postal codes with standard geographic areas should also be more rigorously tested by comparing the 2001 Census results with that of the 2006 Census.

The 2006 Census reported postal codes can be used as follows for geocoding:

- For PCTypes 1 and 2 that service particular sets of dwellings or buildings the intention would be to geocode to the block-face address range. The FSA service area can be constructed with reasonable confidence. Records from the Address Lookup File can be geocoded within these FSA areas. The FSA search areas can be supplemented with search areas created using the relationships between the current census CSD and the previous census CSD. Geocoding can be attempted to the smallest standard geographic area for these by searching for the street name and address range within the service area.
- PCType 4 records do not contain civic address ranges. However, the reporting of these postal codes would indicate where these postal codes are in use, and the same postal codes reported in that area in more than one census would confirm the stable pattern of usage – of well defined routes.

- Service from the postal delivery installation is provided for PCTypes 3 and 5. At first glance it would seem that these should only be geocoded to the postal delivery installation. However, the census data shows that in rural delivery areas these postal codes are commonly used and may be the only postal codes available. In rural delivery areas PCType 5 can be assigned based on the census reporting. In rare instances where there is no postal code reported for a populated rural area, it is likely that when a

18. The method being used for geocoding postal codes to the 2001 Census geographic areas is based on matching CPC community name to CSD name. This method has problems as discussed in section 1.3. However, improvements are being made to the matching to CSD name and this is discussed further in Appendix C.
19. Where large-scale changes do occasionally affect FSA areas, these changes are reported by CPC.

postal code is used, the postal code used is that of the nearest postal delivery installation.

Geocoding directly to standard geographic areas from the census data rather than converting previously geocoded data should result in geocoding that is consistent with the concept of postal code service areas as well as more complete coverage and more precise geocoding.

3.0 Geocoding rules and measures of quality

Rules based geocoding can be developed by combining the information from the census and the CPC files. Quality indicators can be developed for each record geocoded with this approach since both the rules based method of geocoding and the information used for geocoding can be used to create the quality indicators.

3.1 Service areas and search areas

Service areas need to be defined to start the geocoding process. The service areas and relationships suggested for geocoding are:

1. FSA service area (census-based)
2. Postal code service area (census-based)
3. CPC community to CSD correspondence file

FSA service area, as discussed earlier, can be used to find and geocode to civic address ranges. For the 2006 Census, each FSA service area would include every DA which had at least one dissemination block where a minimum number of respondents reported that FSA in the census.[20] Having a minimum number of respondents would ensure that the FSA is truly used and reported accurately within the DA. Because more than one FSA may be used within a DA, FSA boundaries constructed of whole DAs will overlap. However, a little overlap and overcoverage should not be an issue since these FSA service areas will not be used to geocode directly. FSA service areas will only provide the search areas for finding and geocoding postal codes based on street address information in the Road Network Files.

The postal code service area, unlike the FSA service area will be more tightly constructed to include only dissemination blocks where this postal code was reported by a minimum number of respondents. Furthermore, in order to restrict overlaps, only the highest reported postal code in that dissemination block would be chosen. The tighter delineation criteria is used because the postal code service area may be used to directly geocode the postal code to a dissemination block where the postal code usage was reported.

However, not all postal codes may be assigned a search area based on the 2006 Census response data.[21] Imputed postal codes could be used to create more comprehensive search areas but, this is generally a little less reliable. Another possibility is to create the relationship between the CSD used for the census with that of the previous names of the CSD as well as the place name file maintained in Geography Division.

---

20. For the test of FSA service areas (mentioned in section 2.2) a 2001 Census block was assigned to a FSA *service area* only if a minimum of four respondents reported a postal codes with that FSA in that block. Choosing a minimum number of responses of an FSA is based on the premise that the likelihood that four people making the same mistake in reporting the FSA is lower that that of one person making that mistake in the same block. Despite using this conservative criterion, the geocoding rate was over 88% of the PCType 1 and 2 civic address records.
21. Based on the test of FSA service areas mentioned above, *service areas* could not be constructed for all of the postal codes because not all postal codes were reported.

The CPC community's coverage indicates the area where the street name is unique. (As discussed earlier, CPC retains the old municipality names in cases of amalgamations and boundary changes until the duplicate street address problems are resolved.) Then this CPC community can match to one or more CSDs may be inexact. It is difficult to say definitively where the boundaries of the CPC community are because these boundaries are not available from CPC. However, CSDs can be matched to the CPC community. This match between the CSDs and the CPC community can be used as confirmatory information for the automated geocoding by FSA service area.[22] However, because the correspondence is not exact in terms of area covered, this correspondence will not be used as the primary method to create search areas for automated geocoding. The CPC community to CSD correspondence will only be used as an additional check on the automated geocoding and as a secondary means of geocoding.

3. 2 Geocoding postal codes by PCType

Given the intention to geocode the postal codes as definitively as possible, priority will be given to geocoding postal code records whose service areas are defined by civic address ranges. These are the PCTypes 1 and 2. The next most definitive coverage is likely PCType 4 since the routes serviced are defined even if they do not have civic addresses or road names that can be matched to a road network file. Lastly, PCTypes 3 and 5, which are associated with the postal delivery installation but, can be accessed by households in various parts of the CPC community or even by travellers or others outside the municipality will be given priority in terms of the allocation of the standard geographic areas to the postal code. This allocation of standard geographic areas in terms of the type of service available there puts the emphasis on the most likely used postal code in that standard geographic area.

PCTypes 1 and 2 can be definitively geocoded by matching the CPC address information to the Road Network File. If the match is attempted on the street network and address information alone, there can be potentially multiple matches (across Canada). For example, there are many "Main" streets in Canada. Searching on a street name of "Main" would result in multiple matches. In the automated matching program, the matches would be done within a search area. The testing done to date suggests that the FSA search area should be used as the primary search area. The CSD of the match can also be confirmed by the CPC community to CSD Correspondence File. (Please see Appendix C on creating a correspondence between CPC community and CSD). PCType 1 and 2 records should be geocoded to the block-face or dissemination block level since these clearly represent postal delivery to address ranges. The precision of the geocoding can also be monitored – the number of dissemination blocks or block-faces geocoded to should be proportional to the address range records from the CPC Address Lookup File. In this way, both the census and the CPC Address Lookup File information can be combined to do the geocoding and assess its quality.

Various combinations of PCTypes 2, 3, 4 and 5 occur together, particularly in rural delivery areas. (See table 5.) The census reported postal codes do not contain information on which PCType record was reported. Therefore, the type of service from that postal code can not be deduced. The Address Lookup File from CPC does provide the civic address information for PCType 2 but the coverage of each of the other PCTypes for a given postal code can not be deduced. Given this, the geocoding of the share of a particular type of postal code of PCTypes 3, 4 or 5 may not be accurate in terms of usage. In the past, postal codes associated with installations, particularly of PCType 3, were expected to be coded (on the PCCF) to the dissemination block of the installation. The approach here is to continue with that coding while distinguishing the record on the PCCF as one referring to the installation. But, where the general delivery (PCType 5) and post office box (PCType 3) are the only service available, these could also be geocoded to the

22. A list of correspondence between the *CPC community* names and the CSDs has been created for searching of street addresses in the manual geocoding process. Please see Appendix C for more details on how this correspondence file is being improved.

standard geographic area as the postal code most likely in use there. The following approach is suggested for postal codes with any permutation of PCTypes 2, 3, 4 and 5 records:

1. If the postal code has PCType 2 records, within the FSA search area, code all the PCType 2 records based on address information from the Address Lookup File.
2. If the *postal code* service area dissemination blocks have not been completely geocoded as a result of the above step, assign remaining incompletely geocoded (or not geocoded) dissemination blocks and DAs of that *postal code* service area as possible to the PCType 4 records. Where the postal codes are associated with only PCType 4 or only PCTypes 4 and 2 records *all* remaining records at this point can be allocated to PCType 4. However, where PCType 4 occurs with PCTypes 3 and/or 5 (less than 0.5 % of the postal codes) manual resolutions with maps may be needed to ensure proper coverage of the routes associated with the PCType 4 records.
3. Any remaining unassigned standard geographic areas of the *postal code* service area could be assigned to the PCType 5 records at the DA level on the basis that this service could be accessed from the postal delivery installation for households in this area. Similarly, where the general delivery (PCType 5) service is not available, the dissemination blocks could be coded to the PCType 3 of that postal code service area.

In sparsely populated DAs where the postal service used is not evident from the response data or based on address information, the service available is likely from the nearest postal installation. These DAs could be coded to the postal code of the nearest postal delivery installation where postal service could be accessed (PCType 5 or 3). Geocoding done in this manner could be flagged using a field indicating the source of geocoding. The intent is to provide complete geocoding of postal codes in terms of the service available in every populated DA.

The method of geocoding described here uses the FSA service area based on the census as the primary search area. PCType 1 and 2 are then geocoded within this area by matching on street address information. Where this is not possible, the relationship between the CPC community and FSA to the CSD could be used. The geocoding of PCType 3, 4 and 5 records is more difficult since these do not map to civic addresses. However, the census responses can be used to code to areas where postal codes of this type are in use. PCType 3 and 5 would also continue to be coded to the postal installation as well as the service area. In this way, all postal codes in use can be mapped to geographic areas based on the available census response data and the Road Network File.

3.3 Measures of quality

The quality of the postal code geocoding can be both improved and better reported given the concept of geocoding to FSA and postal code service areas. (Background and definitions of the measures of quality are provided in Appendix B).

A. Timeliness

Timeliness can be measured in terms of the difference between the following:

1. The vintage of the postal codes and the release date of the postal products;
2. The vintage of the census and the release of the postal code products.

The conceptual review above allows for rules based geocoding of postal codes to block-faces, dissemination blocks and DAs in Statistics Canada's Geographic Frame. Automated geocoding has been tested and can improve the efficiency of the geocoding process. While it may not be possible to use automated geocoding when there is insufficient information on the Geographic Frame or inconsistencies between the Geographic Frame and CPC data, it can be used for the majority of records. This efficiency can have the following improvements in vintage:

- Fewer manual geocoding resources need to be spent to make the PCCF and PCFR available;

- The time between the availability of the 2006 Census results and the release of the postal code products (PCCF and PCFR) can be reduced.

There are vintage differences between the census response data (available once in five years), the files from Canada Post (available once a month) and the Road Network Files (which are periodically updated). These can lead to difficulties in creating the association or linkage in the geocoding process. The lack of up-to-date information in the Road Network Files is solved usually by contacting the appropriate municipality and requesting information. This can be a time-consuming process and can also result in response burden to the municipality. One way to deal with this would be to stage the geocoding. That is, code to only the CSD or the DA if the road network information is not available and then code to more precise levels as the information becomes available. Currently, once a postal code is geocoded we do not go back and try to improve the precision. In fact we may never re-geocode the postal code unless Canada Post signals a change in the service for that postal code. With automated geocoding we would have the ability to look for improvements in precision with each iteration of the Road Network File.

B. Accuracy and precision

The accuracy of the geocoding process can be indicated in terms of the confidence of geocoding to the correct standard geographic area(s). Precision can be measured in terms of whether the number of block-faces and dissemination blocks coded to are consistent with the address ranges serviced by the postal code.

A record level data quality rating can be created based on the geocoding process to indicate the confidence rank of the geocoding. Access to more than one source of information for the geocoding also allows for the construction of quality indicators for the geocoding. One way of assigning the quality indicator would be to construct a three character indicator based on the geocoding process itself. The characters would be easily understood with "AAA" suggesting that the quality of the geocoding was confirmed by more than one source and was judged as the best and with "CCC" indicating that the geocoding was an approximation. The following table illustrates the proposed concept:

Quality indicator for PCType 1 and 2

The final Quality indicator (QI) is constructed as a concatenation such that:

QI = QI_1 I QI_2 I QI_3

A QI = AAA indicates that the geocoding was of the best quality.

The quality indicator QI_1

This indicator is for the quality of the general area to which the geocoding was done. It should be an indicator of the certainty in terms of geocoding to the correct census subdivision (and correct part of that CSD).

A Search area is based on FSA service area and CPC community to CSD correspondence
B Search area is based on FSA service area only
C Search area is based on CPC community to CSD correspondence only

The quality indicator QI_2

This is an indicator of the street matching for PCTypes 1 and 2.

A Match is found on standardized street name, type and direction to the Road Network File
Match is found on street name and type to only one street and the street has no direction associated with it on the Address Lookup File or the Road Network File
B There is only one street and type combination that matches within search area
C Satisfactory match is found on street name only or street name and direction

The quality indicator QI_3

QI was assigned as follows in matching address ranges to the Road Network File.

A The parity was matched on both addresses to the Road Network File
B The parity was matched on one address only to the Road Network File
C The parity was not matched but the ranges overlap on the Road Network File

This quality indicator applies to the street address matching process for PCType 1 and 2 (about 97% of the records on the Address Lookup File). A data quality indicator could also be provided in a similar manner for postal codes coded directly to an installation.

Where the geocoding is done based on the census information alone, a measure of stability could be created by comparing the geographic area reported in the previous census. This validation is not a direct measure of accuracy but, does provide an indicator of the confidence of the geocoding.

Quality indicator for PCType 3, 4, and 5

The final quality indicator (QI) is constructed as a concatenation such that:

QI = QI_1 | QI_2

A QI = AA indicates that the geocoding was of the best quality. There is no QI_3 available for PCType 3,4 and 5 since there are no street address range information describing the service area and the geocoding is not done to the same precision.

The quality indicator QI_1

This indicator is for the quality of the general area to which the geocoding was done. It should be an indicator of the certainty in terms of geocoding to the correct census subdivision (and correct part of that CSD.) QI_1 is constructed in exactly the same way as for PCTypes 1 and 2.

A Search area is based on FSA service area and CPC community to CSD correspondence
B Search area is based on FSA service area only
C Search area is based on CPC community to CSD correspondence only

The quality indicator QI_2

This is an indicator of the stability of the definition of the boundaries of these service areas. QI_2 is constructed differently for PCType 3, 4 and 5 (from that for PCType 1 and 2) since no street address range information is available for the geocoding.

PCTypes 3, 4, 5 (Postal code service area location)

A Stable area (i.e., geocoded to the same DA) based on current census data and previous census data
B Area is based on current census data and verified via telephone with the municipality
C Area is based on current census data alone

PCType 3 (Postal installation location)

A Address from Householder File is geocoded to the Road Network File and matches the DA of the point file from CPC[23]
B Address from Householder File is geocoded to the Road Network File
C Postal installation located based on the point file from CPC alone

Precision of the area coded to can be measured in terms of the level of geography coded to as well as the number of standard geographic areas coded. This is also particularly relevant for the civic address based geocoding. For example, PCTypes 1 and 2 have service areas defined by address ranges. This would imply that at the most precise level the geocoding would be done to a block-face, followed by the dissemination block and then the DA, the DA being the least precise level of geocoding. The number of dissemination blocks geocoded to as a proportion of the number of address ranges for a postal code would also indicate whether there has been allocation of unnecessary dissemination blocks to a postal code. Measures of precision can be used as part of the verification of the geocoding process as well as to provide general indicators to the data user.

23. A file of delivery installation locations was obtained from Canada Post in January 2006 for testing.

C. Completeness

Given that the postal code products (PCCF and PCFRF) provide links between postal codes and standard geographic area, the completeness of the coverage of these products can be measured in the following way:

1. The degree to which all postal codes and their PCTypes are coded to standard geographic areas;
2. The degree to which every populated standard geographic area is assigned a postal code.

The change in methodology proposed here allows for postal codes of PCType 3 and 5 to be coded to populated DAs when no postal walk or route service is available. This would result in a more complete coverage of the postal service available across Canada.

D. Relevance

The sales as well as the use within Statistics Canada of the PCCF and the PCFRF show that this product is useful. However, an indicator in the file, the Single Link Indicator (SLI) is used for population or dwelling allocation although it is not truly designed for that purpose. The SLI is designed to provide one representative point for a postal code for the purpose of mapping label points of the postal code. But, this indicator is also being used for data allocation. One standard geographic area (dissemination block or DA) may not be representative of all of the population accessing this postal code. For example a postal code service area may span three dissemination blocks and one dissemination block may not be more representative than the others. The recommendation is that data allocation be done by either using statistical software available from Statistics Canada (PCCF+) or by using another allocation process based on a DA population weighted file.[24]

E. Interpretability and metadata

Metadata can be provided at the record level to enable clients to better understand the record linkage. These metadata will also help in the internal evaluation of the improvements that may be needed in geocoding. The quality indicator (QI) has been discussed earlier (in section 3.3). The other record level metadata suggested are described below:

- Method of coding

The method of coding can be relevant to the quality of the record linkage. The following are proposed for this indicator:

1. Manual coding confirmed with maps;
2. Manual coding confirmed with call to relevant municipal or provincial authority;
3. Automated geocoding;
4. Geocoding done to 2001 geographic area re-based to 2006 geographic area.

Currently the QI is proposed to be available only for the automatically geocoded records. This additional indicator documents the source of the other types of record linkage. With the monitoring of the quality of these other types of coding, the QI could in the future be made to include all sources of coding.

---

24. The Postal Code Conversion File Plus (PCCF+) is a complementary product to the Postal Code Conversion File (PCCF). It is an automated system that uses postal codes to assign census geography. PCCF+ is based on the latest Postal Code Conversion File and the Postal Code Population Weight File produced by the Geography Division of Statistics Canada. PCCF+ uses weights to allocate postal codes linked to multiple dissemination areas according to the distribution of population using a given postal code.

- Service areas coded

The postal code may be coded to the postal code service area or to a postal delivery installation and these two types of coding need to be distinguished. The following is proposed for this indicator:

1. Coded to postal code service area;
2. Coded to postal delivery installation.

- DMT with modifications

Currently where the DMT is missing, Statistics Canada assigns a "W" for the DMT. This "W" includes both rural and small urban areas. The user may be able to better use the DMT if the distinction can be made between these. Another letter such as "S" could be used to distinguish rural delivery areas where the PCType = 2 and delivery to a civic address is available.

Geocoding rules and well defined sources of the geocoding enable the creation of quality indicators and other metadata that can inform the user as well as enable the monitoring of the quality of the files.

Conclusion

The intentions of this paper were: (1) to study the postal code concepts and descriptors used by Canada Post Corporation; (2) to establish a conceptual framework articulating the relationship between postal codes and standard geographic areas; and, (3) to subsequently, establish business rules for linking postal codes to standard geographic areas.

Postal codes can be related to standard geographic areas by the postal code service areas. While in some cases the address information available from Canada Post is sufficient to *geocode* the postal code to a road network file, in many cases supplementary information is needed to do the geocoding. The census provides an ideal means to discern the service area of the FSA and also that of the postal code. While this method of geocoding emphasizes the household use of the postal code, business postal codes can also be geocoded this way.[25]

For the geocoding of the postal code, the first step would be determining the FSA service area based on the census and supplementing this with a concordance file of the CPC community and the CSD. This would establish the search area within which the postal code *can be* geocoded. Differentiating the geographic footprint of the postal code by PCType enables better coding to postal code service areas. An automated geocoding process based on this methodology would allow for efficiency in geocoding. An efficient geocoding system can be run as needed to take advantage of updates to postal codes as well as improvements in the road network. Ultimately, the conceptual framework discussed here, provides rules for an efficient geocoding process with metadata on the quality of the geocoded record.

25. The authors suggest that the business use of the postal code as well as comparisons of the postal code service area footprint between the 2001 Census and the 2006 Census can be explored in a subsequent working paper.

Appendix A - Postal code structure

The postal code is an alpha-numeric combination of six characters describing the destination of each item of mail addressed in Canada. The characters are arranged in the form "ANA NAN" where "A" represents an alphabetic character and "N" represents a numeric character (e.g. K1A 0T6). The postal code uses 20 alphabetic characters and 10 numeric characters. Six alphabetic characters (D, F, I, O, Q, and U) are not in use at the present time.

The first character of a postal code is allocated in alphabetic sequence from east to west across Canada and denotes a province, territory, or a major sector found entirely within the boundaries of a province.

Province/Territory/Region	First character of postal code
Newfoundland and Labrador	A
Nova Scotia	B
Prince Edward Island	C
New Brunswick	E
Quebec East	G
Montreal Metropolitan	H
Quebec West	J
Eastern Ontario	K
Central Ontario	L
Toronto	M
South-western Ontario	N
Northern Ontario	P
Manitoba	R
Saskatchewan	S
Alberta	T
British Columbia	V
Northwest Territories and Nunavut	X
Yukon Territory	Y

The first three characters of the postal code ("ANA") represent a set of well-defined and stable areas known as forward sortation areas (FSAs). The FSA represents a specific area within a major geographical region or province/territory.

Appendix B – Information data quality and spatial data quality

The essence of Statistics Canada's policy on data quality and methodology is that users must be provided with the information necessary to judge its fitness for their intended use. Six dimensions of quality identified within the concept of "fitness for use" are shown in the table below:

The six dimensions of information quality

1. The ***relevance*** of statistical information reflects the degree to which it meets the real needs of users. It is concerned with whether the available information sheds light on the issues of most importance to users. The assessment of relevance needs to take into account the varying needs of users.
2. The ***accuracy*** of statistical information is the degree to which the information correctly describes the phenomena it was designed to measure. It is usually characterized in terms of error in statistical estimates and is traditionally decomposed into bias (systematic error) and variance (random error) components. It may also be described in terms of the major sources of error that potentially cause inaccuracy (e.g., coverage, sampling, nonresponse, response).
3. The ***timeliness*** of statistical information refers to the delay between the reference point (or the end of the reference period) to which the information pertains, and the date on which the information becomes available. It is typically involved in a tradeoff against *accuracy*. The *timeliness* of information will influence its *relevance*.
4. The ***accessibility*** of statistical information refers to the ease with which it can be obtained by users. This includes the ease with which the existence of information can be ascertained, as well as the suitability of the form or medium through which the information can be accessed. The cost of the information may also be an aspect of *accessibility* for some users.
5. The ***interpretability*** of statistical information reflects the availability of the supplementary information and metadata necessary to interpret and utilize it appropriately. This information normally covers the underlying concepts, variables and classifications used, the methodology of collection, and indicators of the accuracy of the statistical information.
6. The ***coherence*** of statistical information reflects the degree to which it can be successfully brought together with other statistical information within a broad analytic framework and over time. The use of standard concepts, classifications and target populations promotes coherence, as does the use of common methodology across surveys. *Coherence* does not necessarily imply full numerical consistency.

Spatial data quality elements provide information on the fitness-for-use of a spatial data set by describing when and how the data were created, the consistency of the data set with others and the accuracy of the data. The elements include an overview describing the purpose and usages, as well as specific quality elements reporting on the lineage, positional accuracy, attribute accuracy, logical consistency and completeness. The elements of spatial data quality are listed as follows:

Elements of spatial data quality

1. **Lineage**: Describes the history of the spatial data, including descriptions of the source material from which the data were derived, and the methods of derivation. It also contains the dates of the source material, and all transformations involved in producing the final digital files or map products.

2. **Positional accuracy**: Refers to the absolute and relative accuracy of the positions of geographic features. Absolute accuracy is the closeness of the coordinate values in a dataset to values accepted as or being true. Relative accuracy is the closeness of the relative positions of features to their respective relative positions accepted as or being true. Descriptions of positional accuracy include the quality of the final file or product after all transformations.

3. **Attribute accuracy**: Refers to the accuracy of the quantitative and qualitative information attached to each feature (such as population for an urban area, street name, census subdivision name and code).

4. **Logical consistency**: Describes the fidelity of relationships encoded in the data structure of the digital spatial data.

5. **Completeness**: Refers to the degree to which geographic features, their attributes and their relationships are included or omitted in a dataset. It also includes information on selection criteria, definitions used, and other relevant mapping rules.

Appendix C – CPC community and CSD

The primary approach to geocoding that is recommended in this paper is using service areas to create search areas. However, historically the primary approach to geocoding has been to locate the CSD or part of the CSD represented by the CPC community and then geocode based on the address information. The correspondence between the CPC community and the CSD needs to be built on more than just a similarity in names. This approach, of geocoding by finding the CSD is recommended now as a secondary approach. It would allow for geocoding where it was not possible to create service areas or to confirm geocoding done using the FSA service areas.

The manual approach to geocoding postal codes in Statistics Canada has been to start with matching the CPC community name from the Address Lookup File to information from a road network file. This was usually done by trying to match the CPC community name to the CSD name. Where this was not possible, supplementary maps were consulted and phone calls were made to municipal authorities to aid in this process. As discussed in this working paper (in section 1.3.1) the CPC community names only directly match the CSD names less than 40% of the time and this is part of what makes the geocoding a time consuming and inexact process. Furthermore, even where the match is found for the name, the area implied by a CPC community may be different from the boundaries of the CSD by the same name.

Improving the match between the CPC community and the C[illegible] would provide supplementary information for geocoding. These improvements have been done in the following way:

1. Matching CPC community name directly to CSD name after standardization of spelling ;
2. Matching CPC community name to a previous census CSD name and then finding the related current census CSD;
3. Matching the CPC community name to a place name and then finding the related current census CSD.

Because CPC often continues to use old municipality names, this approach allows for the match of historic names to the current CSD. This matching has proved to be helpful in the manual geocoding process.

However, this approach could lead to potentially geocoding to two different streets having the same name in the CSD. (CPC maintains the older municipality name, generally until issues related to such duplicate names are resolved). Because of this, an approach of matching older CSD boundaries to current geographic areas (such as the dissemination area or dissemination block) would perhaps provide a better search area for geocoding. This approach needs to be tested to see whether search areas created in this way can be used for automated geocoding.

Because a match of CPC community to CSD, based on name alone, may be inexact in terms of area, this is not recommended as the primary means of geocoding. A match of the CPC community's area to lower level current census *geographic areas* could provide a search area that could be used for automated geocoding. Testing is needed to see how well this method of geocoding can supplement the geocoding based on service areas. The manual geocoding process is currently being done independent of the work on automated geocoding. The improvements in matching the CPC community to the CSD have already been found effective in the manual geocoding process.

Glossary

Block-face
A block-face is one side of a street between two consecutive features intersecting that street. The features can be other streets or boundaries of standard geographic areas. Block-faces are used for generating block-face representative points, which in turn are used for geocoding and census data extraction when the street and address information are available

Civic address
Postal addresses can be categorized as civic addresses or postal installation addresses. A civic address consists of street address number, a street name, municipality name and postal code. The street name may include a street type and direction. The street address number may include a suite and/or apartment number (Canada Post Corporation, 2005a, p. 20-21).

Census subdivision (CSD)
Census subdivision (CSD) is the general term for municipalities (as determined by provincial legislation) or areas treated as municipal equivalents for statistical purposes (for example, Indian reserves, Indian settlements and unorganized territories).

Census subdivision type
Census subdivisions (CSDs) are classified into 55 types according to official designations adopted by provincial/territorial or federal authorities. Two exceptions are "Subdivision of unorganized (SNO)" in Newfoundland and Labrador, and "Subdivision of county municipality (SC)" in Nova Scotia, which are geographic areas created as equivalents for municipalities by Statistics Canada, in cooperation with those provinces, for the purpose of disseminating statistical data. The census subdivision type accompanies the census subdivision name in order to distinguish CSDs from each other, for example, Granby, V (for the *ville* of Granby) and Granby, CT (for the *municipalité de canton* of Granby).

CPC community
The term "municipality" as in the Address Lookup File does not necessarily constitute the legally defined municipality at the time when Canada Post publishes its data. This term serves the purpose of preserving locally used names to prevent confusion in reference to the area serviced. To avoid confusion this term is referred to in this paper as the "CPC community".

Delivery Installation
This is a local facility from which mail is delivered. Post offices, stations, retail postal Outlets, centre postal communitaires, and letter carrier depots are different categories of delivery installations (Canada Post Corporation, 2005a, p. 22-23).

Dissemination area (DA)
A dissemination area (DA) is a small, relatively stable geographic unit composed of one or more adjacent dissemination blocks. It is the smallest standard geographic area for which all census data are disseminated. DAs cover all the territory of Canada.

Dissemination block
A dissemination block (DB) is an area bounded on all sides by roads and/or boundaries of standard geographic areas. The dissemination block is the smallest geographic area for which population and dwelling counts are disseminated. Dissemination blocks cover all the territory of Canada.

Dwelling
In general terms a dwelling is defined as a set of living quarters. Two types of dwelling are identified in the census: collective dwellings and private dwellings. The former pertains to dwellings which are institutional, communal or commercial in nature. The latter, private dwelling refers to a separate set of living quarters with a private entrance either from outside the building or from a common hall, lobby, vestibule or stairway inside the building. The entrance to the dwelling must be one that can be used without passing through the living quarters of some other person or group of persons.

Forward sortation area (FSA)
The first three characters of the postal code (alphanumeric combination of "ANA") form the forward sortation area (FSA). This is defined quite generally on the CPC internet site as follows:

> The forward sortation area or (FSA) represents a specific area within a major geographic region or province. The forward sortation area provides the basis for the primary sorting or forward mail *(Canada Post Corporation, 2006, Section B, Chapter 3, p. 9)*.

FSA service area
This is defined in this paper as the area where persons have access to mail for postal codes with the same foreword sortation area (FSA).

General delivery (GD)
A service provided at postal outlets for delivery of mail to those not renting post office boxes, the traveling public, customers with no fixed address within the letter carrier delivery area and anyone who cannot receive mail through the normal delivery modes (Canada Post Corporation, 2007, Glossary).

Geocoding
Geocoding is the process of assigning geographic identifiers (codes) to map features and data records. The resulting geocodes permit data to be linked geographically.

Household
The concept of Household applies to a person or group of persons who occupy the same dwelling and do not have a usual place of residence elsewhere in Canada or abroad. The dwelling may be either a collective dwelling or a private dwelling.

Lock box
Lock boxes are usually set in post office walls in groups. Mail is sorted by individual box number from the employee working area. The customer removes it after unlocking the door on the public side with the key supplied when the box is rented (Canada Post Corporation, 2007, Glossary).

Mobile service
Mobile service is delivery of mail to industrial areas by motorized means (usually performed under contract), (Canada Post Corporation, 2007, Glossary).

Post office box
See lock box

Postal code service area
This is defined in this paper as the area where persons have access to mail for that postal code.

Postal installation address (or "Station" address)
Postal addresses can be categorized as civic addresses or postal installation addresses. A postal installation address consists of a type of delivery, which may be general delivery, lock box number or (rural route number, mobile route number or suburban service number)., municipality name (which is referred to here as the CPC community name), province and postal code. The address may also include the CPC *delivery installation* information (Canada Post Corporation, 2005a, p. 20-21).

Rural delivery area
This term is used to define areas where service is provided by postal codes defined as "rural" by CPC. The CPC classification is considered to be a component of mail preparation. This classification is typed based on the 2nd character of the postal code as follows:

- an urban postal code: numerals 1 to 9 (E2J). Urban postal codes are generally serviced by letter carrier or community mailboxes; or
- a rural postal code: numeral 0 (zero) (A0A). Rural postal codes are serviced by rural route drivers and/or postal outlets.

Rural delivery areas do not represent the same area as Statistics Canada's rural areas.

Rural route (RR)
This is a system for collection and delivery of mail and for the transaction of other postal business with customers residing along or near well defined roads in reasonably well settled rural areas (Canada Post Corporation, 2007, Glossary).

Search area
The term search area is used in this paper to denote the geographic area within which the geocoding can be done using road names and civic addresses.

Single Link Indicator
The single link indicator can be used to force a one-to-one relationship between postal codes and dissemination areas, dissemination blocks or block-faces. This relationship may be useful for mapping, however, because many dissemination areas, dissemination blocks or block-faces are not represented by the SLI, this is not recommended for uses such as allocation of population to geographic areas.

Standard geographic areas
In the working paper, standard geographic areas refers to the geographic areas used for the publishing of census data including dissemination block, dissemination area and census subdivision.

Statistical unit
A Statistical unit is the unit of observation or measurement for which data are collected or derived. The basic statistical unit for social statistics is the person, which is the unit of analysis for most social statistics programs. Other units used in social statistics include the dwelling and the household.

Street address
See civic address.

Urban delivery area

This term is used to define areas where service is provided by postal codes defined as "urban" by CPC. The CPC classification is considered to be a component of mail preparation. This classification is typed based on the 2nd character of the postal code as follows:

- an urban postal code: numerals 1 to 9 (E2J). Urban postal codes are generally serviced by letter carrier or community mailboxes; or
- a rural postal code: numeral 0 (zero) (A0A). Rural postal codes are serviced by rural route drivers and/or postal outlets.

Urban delivery areas do not represent the same area as Statistics Canada's *urban areas*.

References

Canada Post Corporation (2005 a). Postal Code Address Data Technical Specifications, January 2005.
http://www.canadapost.ca/offerings/address_management/can/other_data-e.asp.
(accessed on October 19, 2006)

Canada Post Corporation (2005 b). Householder Data Technical Specifications, January 2005.
http://www.canadapost.ca/offerings/address_management/can/other_data-e.asp
(accessed on October 19, 2006)

Canada Post Corporation (2005 c). Delivery Mode Data Technical Specifications, January 2005.
http://www.canadapost.ca/offerings/address_management/can/other_data-e.asp
(accessed on October 19, 2006)

Canada Post Corporation. 2006. Canada Postal Guide.
http://www.canadapost.ca/tools/pg/manual/b03-e.asp#67044
(accessed on January 15, 2007)

Statistics Canada. 2003. Illustrated Glossary
http://geodepot2.statcan.ca/Diss/Reference/COGG/Index_e.cfm
(accessed on October 19, 2006)

Statistics Canada. Postal Code Conversión File (PCCF), Reference Guide. March 2006.
http://www.statcan.ca/bsolc/english/bsolc?catno=92F0153GIE
(accessed on October 19, 2006)

Statistics Canada. Policy on Informing Users of Data Quality and Methodology. Approved March 31, 2000.
http://www.statcan.ca/english/about/policy/infousers.htm
(accessed on October 24, 2006).